Simple Varieties for Empirical Points

Claudia Fassino[*] Maria-Laura Torrente[*]

Abstract

Abstract

We present a symbolic-numeric approach for the analysis of a given set of noisy data, represented as a finite set $\mathbb{X}$ of limited precision points. Starting from $\mathbb{X}$ and a permitted tolerance ε on its coordinates, our method automatically determines a low degree monic polynomial whose associated variety passes close to each point of $\mathbb{X}$ by less than the given tolerance ε.

1   Introduction

It is a well-known matter that the analysis and modeling of real-world phenomena often relies upon measurements and observations which give rise to sets of data represented as affine sets of points. An algebraic way that exploits the collected data to construct a mathematical model of the observed phenomenon consists in computing the vanishing ideal of the affine set of points, that is the ideal comprising all polynomials which vanish at the given points. The vanishing ideal is classically determined using the Buchberger-Möller (BM) Algorithm [5], a low complexity method which returns a Gröbner basis of it. Nevertheless, the fact that data and relations from the real world are essentially characterized by a limited accuracy makes this algebraic approach

---

[*]Dip. di Matematica, Università di Genova, via Dodecaneso 35, 16146 Genova, Italy (`fassino@dima.unige.it, torrente@dima.unige.it`)

unfeasible in practice. In fact, even small perturbations of the points end up with very different vanishing ideals, which may have completely different bases, since the question whether or not a polynomial vanishes at a given set of points is obviously very sensitive to perturbations. For instance, this is a well-known phenomenon when using Gröbner basis theory ([17], [18]) which turns out to be unsuitable as a numerical tool. Thus, when dealing with real-world phenomena, the sole use of the classical algebraic techniques turns out to be insufficient, and a combination with numerical tools is required.

Motivated by the need of describing real-world processes, the development of a numerical branch within a discrete mathematical discipline already took place in different areas of mathematics, such as Linear Algebra, Differential Equations, Optimization, etc. In the nineties, classical nonlinear algebra began to follow this trend, also thanks to the work of H. J. Stetter, whose book [21] is considered as a stepping stone at the interface of symbolic and numerical computation. This new emerging discipline has been given different names (see also [19]); in [22] its mathematical content has been clearly described as including the "areas of commutative algebra not over arbitrary fields or rings but over the analytically structured fields of the real or complex numbers", in which the real or complex metric is taken into consideration. Obviously, in the attempt of generalizing the notions of classical Commutative Algebra to the real or complex coefficients, the greatest attention must be paid regarding the new concepts and terminology. In particular, in our case, the classical notion of vanishing ideal turns out to be inadequate for limited precision points, and so it needs to be replaced by an appropriate concept also suitable for numerical computations.

In the literature a notion of vanishing ideal of a set of limited precision points already exists; in [2], [10], [13], [15], [20], [23] it is defined in different ways as an ideal containing polynomials whose evaluations at the original points assume small values. Note that, according to this definition, the vanishing ideal of a set of limited precision points is not necessarily zero-dimensional; further, maximum relevance is given to the evaluations of the polynomials at the points, while the geometrical distance of the zero-locus of the polynomials from the original points is not taken into consideration. In this paper we follow a different approach: we give more relevance to the distance points-variety and introduce the idea that the vanishing ideal of a set of limited precision points is generated by polynomials of low degree whose affine variety lies close to the original points. Among such polynomials, a crucial role is played by those of minimal degree, which provide an immedi-

ate interpretation of the phenomenon encoded by $\mathbb{X}$ and a simple geometrical representation of the data points, as shown in the following example.

Example 1.1. Let $\mathbb{X} \subset \mathbb{R}^2$ be a set of points created by perturbing by less than 0.1 the coordinates of 10 points lying on the affine variety $g = 0$ where $g = y^2 - x - 2y + 2$:

$$\begin{aligned}
\mathbb{X} \;=\; &\{(0.95, 1), (1.3, 0.5), (2.05, 1.98), (2.08, 0), (3.18, -0.48), \\
&(5.05, 2.95), (5.05, -0.95), (7.2, -1.45), (9.98, 4), (10.05, -2)\}
\end{aligned}$$

Using standard techniques of Computational Algebra we obtain that the minimal degree of the polynomials vanishing at $\mathbb{X}$ is 4; this degree is too high to point out that the set $\mathbb{X}$ lies close to a parabola, while g and its zero-locus $\mathcal{Z}(g)$, almost crossing each point of $\mathbb{X}$, provide a good numerical representation of the given data set.

This paper, which is the first part of a wider investigation, reports on the problem of computing a polynomial f of low degree whose associated affine variety $\mathcal{Z}(f)$, which is defined as the set of points at which the polynomial f vanishes, almost crosses each element of the data set $\mathbb{X}$. The final aim of our future work is to detect a set of polynomials whose minimal degree is strictly bounded by the minimal degree of the elements of the vanishing ideal of $\mathbb{X}$ and such that their zero-set is made up of points each differing from the corresponding point of $\mathbb{X}$ by less than the given tolerance. Put differently, we think to the *numerical* vanishing ideal as a truly zero-dimensional ideal whose *exact* zeros are small perturbations of the original limited precision points, and so numerically indistinguishable from them. An evident merit of this point of view is that the gain in simplicity due to the lower degree of the polynomials might offset the noise in the data and help to discover simpler laws that rule the phenomenon we aim to describe.

In particular, in this paper we address the following problem: given a finite set $\mathbb{X}$ of points and a permitted positive tolerance ε on the coordinates of each point, we look for a polynomial f whose degree is strictly bounded by the minimal degree of the elements of the vanishing ideal $\mathcal{I}(\mathbb{X})$ of $\mathbb{X}$ and whose affine variety $\mathcal{Z}(f)$ lies close to the points of $\mathbb{X}$ by less than ε. We present a new algorithm (the Low-degree Polynomial Algorithm - LPA) that, starting from $\mathbb{X}$ and ε, uses some peculiarities of the SOI algorithm [2] and NBM algorithm [10], both theoretically based on the BM algorithm, to determine a polynomial f of suitable degree. In most cases f also satisfies

3

some simple conditions, derived from Kantorovich theorem [24], under which the variety $\mathcal{Z}(f)$ is proved to lie close to the points of $\mathbb{X}$ by less than ε. In these favourable cases f is a solution of our problem.

More precisely, partly parallelizing the procedures presented in [2] and [10], we reformulate the addressed problem as the problem of solving an ordered finite sequence $\{F_i = 0\}_i$ of nonlinear systems subject to constraints. The first nonlinear system that admits an approximated solution satisfying the constraints allows the direct computation of the polynomial f. In the LPA we solve each nonlinear system $F_i = 0$ using an iterative method which is a modified version of the classical Normal Flow algorithm [24]. Each iteration requires the solution of a linear system whose coefficient matrix is the evaluation of the Jacobian J_{F_i} at a suitable point. Since an analytic expression of F_i as well as of J_{F_i} is computationally very hard to obtain, the evaluation of J_{F_i} is performed using a new method that does not require the explicit knowledge of F_i. Due to the technicalities of this method, its details are separately illustrated inside the Appendix. Moreover, since the evaluation of the Jacobian J_{F_i} often turns out to be an ill-conditioned matrix, the corresponding linear system is solved using a method based on the Rank Revealing decomposition [14].

This paper is organized as follows. In order to be self-consistent and to fix notations, in Section 2 we introduce some basic concepts of Computational Commutative Algebra and Numerical Analysis. In Section 3 we introduce the modified version of the Normal Flow Algorithm to find an approximated solution of an underdetermined nonlinear system subject to constraints, paying particular attention to the ill-conditioned case. Section 4 presents the main results of the paper: the Low-degree Polynomial Algorithm, which performs the main steps of the method described above, and Theorem 4.3, which proves that, under certain hypotheses, the output of the LPA is a solution of our theoretical problem. In Section 5 we give numerical examples illustrating the functioning of the LPA. Finally, 6 contains the basic result for the evaluation of the Jacobian matrix associated with our nonlinear function, without passing through the explicit expression of it.

2 Preliminary definitions and results

We recall some basic notation about matrices. Let $m, n \geq 1$ be integers and $\mathrm{Mat}_{m \times n}(\mathbb{R})$ be the set of $m \times n$ matrices with entries in $\mathbb{R}$; if $m = n$ we

4

simply write $\text{Mat}_n(\mathbb{R})$. Let $A \in \text{Mat}_{m \times n}(\mathbb{R})$; we denote by $\|A\|_p$ the p-norm of A, $p = 1, 2, \infty$.

We divide the rest of this section into three subsections: the first one contains some preliminary definitions about the polynomial ring $\mathbb{R}[x_1, \ldots, x_n]$ and the notion of empirical points; the second one reports some basic results on the numerical rank of a matrix; the third one describes a classical numerical approach to solve underdetermined nonlinear systems.

2.1 Preliminary definitions

We recall some concepts related to the polynomial ring $P = \mathbb{R}[x_1, \ldots, x_n]$ ([16], [17]).

Definition 2.1. Let $\mathbb{X} = \{p_1, \ldots, p_s\}$ be a non-empty finite set of points of $\mathbb{R}^n$, f be a polynomial and $G = \{g_1, \ldots, g_k\}$ be a non-empty finite set of polynomials.

1. $\mathcal{I}(\mathbb{X}) = \{f \in P : f(p_i) = 0 \; \forall p_i \in \mathbb{X}\}$ is the **vanishing ideal** of $\mathbb{X}$.

2. $\mathcal{Z}(f) = \{p \in \mathbb{R}^n : f(p) = 0\}$ is the **affine variety** associated with f.

3. The $\mathbb{R}$-linear map $\text{eval}_{\mathbb{X}} : P \to \mathbb{R}^s$ defined by $\text{eval}_{\mathbb{X}}(f) = (f(p_1), \ldots, f(p_s))$ is called the **evaluation map** associated with $\mathbb{X}$. For brevity, we write $f(\mathbb{X})$ to mean $\text{eval}_{\mathbb{X}}(f)$.

4. The **evaluation matrix** of G associated with $\mathbb{X}$, denoted by $M_G(\mathbb{X}) \in \text{Mat}_{s \times k}(\mathbb{R})$, is defined as having entry (i, j) equal to $g_j(p_i)$, *i.e.* whose columns are the images of the polynomials g_j under the evaluation map.

Let $t = x_1^{\beta_1} \ldots x_n^{\beta_n}$, $\beta_i \in \mathbb{N}$, be a power product and $\mathcal{O}$ be a set of power products; we denote by $\partial_k t = \beta_k x_1^{\beta_1} \ldots x_k^{\beta_k - 1} \ldots x_n^{\beta_n}$ the k-**th formal partial derivative** of t and by $\partial_k \mathcal{O} = \{\partial_k t : t \in \mathcal{O}\}$.

We formalize the idea of perturbed point by introducing a simplified version of the notions of empirical point, admissible perturbation and almost vanishing polynomials (see [1], [2], [10], [21]). To this aim we recall that given a real value $\eta \ll 1$ and a real function $w(x)$ we say that $w(x) = O(\eta^k)$, $k \in \mathbb{N}$, if $|w(x)|/\eta^k$ is bounded near the origin.

Definition 2.2. Let $p, q \in \mathbb{R}^n$ be points and $\varepsilon > 0$.

1. The pair $(p, \varepsilon) = p^\varepsilon$ is called an **empirical point**: p is the **specified value** and ε is the **tolerance** of p^ε.

2. The ε-**neighborhood** of p^ε is defined as $N(p^\varepsilon) = \{\widetilde{p} \in \mathbb{R}^n : \|\widetilde{p} - p\|_\infty \leq \varepsilon\}$ and contains all the **admissible perturbations** of p^ε.

3. Two empirical points p^ε, q^ε are said to be **distinct** if $N(p^\varepsilon) \cap N(q^\varepsilon) = \emptyset$.

Definition 2.3. Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$ be a set of empirical points with uniform tolerance ε and with $\mathbb{X} = \{p_1, \ldots, p_s\} \subset \mathbb{R}^n$

1. A set of points $\widetilde{\mathbb{X}} = \{\widetilde{p_1}, \ldots, \widetilde{p_s}\} \subset \mathbb{R}^n$ is called an **admissible perturbation** of $\mathbb{X}^\varepsilon$ if each $\widetilde{p_i} \in N(p_i^\varepsilon)$.

2. A polynomial g is called **almost vanishing** at $\mathbb{X}^\varepsilon$ if $\|g(\mathbb{X})\|_2/\|g\| = O(\varepsilon)$, where $\|g\|$ is the 2-norm of its coefficient vector.

2.2 The numerical rank

In this section we recall the definition of the numerical rank of a matrix $A \in \mathrm{Mat}_{s \times t}(\mathbb{R})$, and some basic results related to this topic ([11], [14]). In the case $s \geq t$ such results will allow us to detect a partitioning of the columns of A (if $s < t$ the partitioning involves the rows of A) into two submatrices A_1 and A_2 (A_2 eventually the empty matrix) such that A_1 is well-conditioned and A_2 consists of columns (or rows) which are "almost" depending on the columns (or rows) of A_1.

Let $A \in \mathrm{Mat}_{s \times t}(\mathbb{R})$; we denote by $\mathrm{rank}(A)$ the rank of A, by $\sigma_i(A)$ the i-th singular value of A, by $A^\dagger$ the pseudoinverse of A and by $K_2(A) = \|A\|_2 \|A^{-1}\|_2 = \sigma_1(A)/\sigma_{\min\{s,t\}}(A)$ the 2-norm condition number of A. We denote by 0 the vector or the matrix whose elements are equal to 0 and whose dimension is deducible from the context.

Given $k > 1$ and a small threshold δ we say that A has numerical (δ, k)-rank r if there exists a well-determined gap between $\sigma_r(A)$ and $\sigma_{r+1}(A)$ and if both δ and $k\delta$ lie in this gap. This concept can be formalized as follows.

Definition 2.4. Let $r \leq \min\{s, t\}$ and $A \in \mathrm{Mat}_{s \times t}(\mathbb{R})$; let $\delta > 0$ and $k > 1$. The **numerical (δ, k)-rank** of A, denoted by $\mathrm{rank}_{\delta,k}(A)$, is equal to r iff

$$\sigma_1 \geq \ldots \geq \sigma_r(A) > k\delta > \delta > \sigma_{r+1}(A) \geq \ldots \geq \sigma_{\min\{s,t\}}(A)$$

The singular values of A might not clearly split into small and large subsets making the determination of the numerical rank somewhat arbitrary. This leads to more complicated methods for estimating $\text{rank}_{\delta,k}(A)$ which involve the LS problem[11]. In the following of the paper we assume that either the matrix A has full numerical rank or there exists a well-determined gap between $\sigma_r(A)$ and $\sigma_{r+1}(A)$.

Remark 2.5. If $A \in \text{Mat}_{s \times t}(\mathbb{R})$ is a full numerical (δ, k)-rank matrix then it is well-conditioned, that is its condition number cannot be too elevated, as $K_2(A) = \frac{\sigma_1(A)}{\sigma_{\min\{s,t\}}(A)} < \frac{\sigma_1(A)}{k\delta}$. On the contrary, if $\text{rank}_{\delta,k}(A) = r < \min\{s,t\}$ then the matrix is ill-conditioned, that is its condition number is elevated, since $K_2(A) = \frac{\sigma_1(A)}{\sigma_{\min\{s,t\}}(A)} > \frac{\sigma_1(A)}{\delta}$.

The numerical (δ, k)-rank of A points out that there are no matrices with exact rank less than $\text{rank}_{\delta,k}(A)$ which differ from A by less than δ, as the following proposition shows.

Proposition 2.6. *Let $A \in \text{Mat}_{s \times t}(\mathbb{R})$ with $\text{rank}_{\delta,k}(A) = r$.*
Then the set $A_\delta = \{M \in \text{Mat}_{s \times t}(\mathbb{R}) : \|M - A\|_2 < \delta\}$ contains at least one matrix of rank r but no matrix of rank strictly less than r.

Proof. From $\text{rank}_{\delta,k}(A) = r$ it follows that $\sigma_r(A) > \delta > \sigma_{r+1}(A)$; further, from the Eckart-Young theorem [11] for each $j = 1 \ldots \min\{s,t\}$ we have $\sigma_j(A) = \min\{\|A - B\|_2 : \text{rank}(B) = j - 1\}$. Let $C \in \text{Mat}_{s \times t}(\mathbb{R})$ be s.t. $\text{rank}(C) = r$ and $\|A - C\|_2 = \min\{\|A - B\|_2 : \text{rank}(B) = r\}$; since $\sigma_{r+1}(A) < \delta$ then $C \in A_\delta$. Further, since for each $j \le r$ we have $\sigma_j(A) > \delta$; it follows that A_δ contains no matrix of rank strictly less than r. □

The numerical rank of A is preserved under small perturbations, since the following result about the sensitivity of the singular values holds.

Proposition 2.7. *Let $A, E \in \text{Mat}_{s \times t}(\mathbb{R})$; then for each $1 \le j \le \min\{s,t\}$ we have $|\sigma_j(A + E) - \sigma_j(A)| \le \sigma_1(E) = \|E\|_2$*

Proof. See [11], Corollary 8.3.2 □

We end this section with Theorem 2.8 [14] which has important consequences when applied to the case $r = \text{rank}_{\delta,k}(A)$.

Theorem 2.8. *Let $r \leq t \leq s$ and $A \in \mathrm{Mat}_{s \times t}(\mathbb{R})$. There exists a permutation matrix $\Pi \in \mathrm{Mat}_t(\mathbb{R})$ such that*

$$A\Pi = (Q_1 \mid Q_2) \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix} \tag{2.1}$$

where $Q = (Q_1 \mid Q_2)$ is orthonormal, $Q_1 \in \mathrm{Mat}_{s \times r}(\mathbb{R})$, $Q_2 \in \mathrm{Mat}_{s \times t-r}(\mathbb{R})$, $R_{11} \in \mathrm{Mat}_r(\mathbb{R})$ and $R_{22} \in \mathrm{Mat}_{t-r}(\mathbb{R})$ are upper triangular, and

$$\sigma_{\min}(R_{11}) \geq \frac{\sigma_r(A)}{q(t,r)} \quad and \quad \|R_{22}\|_2 \leq q(t,r)\sigma_{r+1}(A) \tag{2.2}$$

where $q(t,r) = \sqrt{r(t-r) + \min(r, t-r)}$.

Proof. See [14], Theorem 2.2. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Theorem 2.8, when applied to the case $r = \mathrm{rank}_{\delta,k}(A)$, states the existence of a set of r strongly independent columns of A or, equivalently, of a well-conditioned submatrix $A_1 \in \mathrm{Mat}_{s \times r}(\mathbb{R})$ of A. In particular, it proves that there exist a permutation matrix Π and a partitioning of $A\Pi = (A_1 \mid A_2)$ where $A_1 = Q_1 R_{11} \in \mathrm{Mat}_{s \times r}(\mathbb{R})$ and $A_2 = Q_1 R_{12} + Q_2 R_{22} \in \mathrm{Mat}_{s \times t-r}(\mathbb{R})$, which has the following interesting numerical properties. If the gap between $\sigma_r(A)$ and $\sigma_{r+1}(A)$ is large enough, that is if $\sigma_r(A) \gg \sigma_{r+1}(A)$, then A_1 is a submatrix of A consisting of the maximum number of strongly independent columns w.r.t. the threshold δ. In fact, from $\sigma_r(A_1) = \sigma_r(R_{11}) > \frac{\sigma_r(A)}{q(t,r)} > 0$, we have that A_1 has full rank, and from $Q_1 = A_1 R_{11}^{-1}$, $\|R_{22}\|_2 \leq q(t,r)\sigma_{r+1}(A)$ and $A_2 = Q_1 R_{12} + Q_2 R_{22}$ we have that each column of the matrix A_2 can be expressed as the sum of a linear combination of the columns of A_1 and a vector whose 2-norm is less than $q(t,r)\sigma_{r+1}(A)$. Furthermore, though A is ill-conditioned (see Remark 2.5), A_1 is well-conditioned, since $K_2(A_1) = \frac{\sigma_1(A_1)}{\sigma_r(A_1)} = \frac{\sigma_1(A_1)}{\sigma_r(R_{11})} \leq \frac{\sigma_1(A)}{\sigma_r(A)}q(t,r)$

2.3 Underdetermined nonlinear systems

Let $D \subseteq \mathbb{R}^n$, $F : D \to \mathbb{R}^m$, with $m < n$, be a differentiable nonlinear function, $J_F(x)$ be the Jacobian matrix of F at x and let $F = 0$ be the underdetermined nonlinear system to solve. In this subsection we recall the Normal Flow Algorithm, a classical iterative method for approximating a solution of $F = 0$ (see [24] and the references given there).

Algorithm 2.9. (The Normal Flow Algorithm - NFA)
Let $D \subseteq \mathbb{R}^n$, $F : D \to \mathbb{R}^m$ be a differentiable nonlinear function, $\omega \ll 1$ be a fixed threshold, and $\bar{x} \in D$ be the initial point. Consider the following sequence of steps.

NF1 Let $h = (1, \ldots, 1)^t$.

NF2 While $\|h\|_2 > \omega$
  - Compute the minimal 2-norm solution of $J_F(\bar{x})h = -F(\bar{x})$.
  - Let $\bar{x} = \bar{x} + h$.

  Return $\bar{x}$ and stop.

At each iteration the minimal 2-norm solution of the underdetermined linear system $J_F(\bar{x})h = -F(\bar{x})$ can be either computed using $J_F^\dagger(\bar{x})$ or, more efficiently, performing a QR decomposition of $J_F(\bar{x})$ (see [9]).

The next classical result (see [24]) is a local convergence theorem for the NFA that also provides sufficient conditions for the existence of a local solution of $F = 0$.

Theorem 2.10. (Kantorovich theorem)
Let $D \subseteq \mathbb{R}^n$ be an open set, $F : D \to \mathbb{R}^m$ be a differentiable nonlinear function and $J_F(x)$ be of full rank m in an open convex set $\Omega \subseteq D$. Let $x_0 \in \Omega$, $\eta > 0$ and $\Omega_\eta = \{x \in \Omega \; : \; \|y - x\|_2 < \eta \Rightarrow y \in \Omega\}$; suppose that

- *$\exists \, \gamma \geq 0$, $p \in (0, 1]$ such that $\|J_F(y) - J_F(x)\|_2 \leq \gamma \|y - x\|_2^p, \; \forall x, y \in \Omega$*

- *$\exists \, \mu > 0$ such that $\|J_F^\dagger(x)\|_2 \leq \mu, \; \forall x \in \Omega$*

Then there exists $\beta > 0$ satisfying $\tau = \frac{\gamma \mu^{1+p} \beta^p}{1+p} < 1$ and $\frac{\mu\beta}{1-\tau} < \eta$ such that if $x_0 \in \Omega_\eta$ and $\|F(x_0)\|_2 < \beta$ then the iterates $\{x_k\}_{k=0,1,\ldots}$ determined by the NFA are well defined and converge to a point $x^* \in \Omega$ such that $F(x^*) = 0$.

Proof. See [24], Theorem 2.1. □

3 Underdetermined and ill-conditioned non-linear systems

As reported in the Introduction, we reformulate the problem addressed in this paper as the problem of solving an ordered finite sequence of underdetermined

nonlinear systems subject to constraints which, in general, turn out to be ill-conditioned; in this section we present a method for solving such kind of nonlinear systems.

Let $D \subseteq \mathbb{R}^n$ be a set containing the origin, let $F : D \to \mathbb{R}^m$ be a nonlinear function of class $\mathcal{C}^2(D)$, and let $Q_\varepsilon = \{x \in \mathbb{R}^n \ : \ \|x\|_\infty \leq \varepsilon\}$; our aim is to compute a solution of the underdetermined nonlinear system $F = 0$ subject to $x \in Q_\varepsilon \cap D$. As reported in Section 2.3 a classical method for finding an approximated solution of $F = 0$ is given by the NFA, which returns successively better approximations of a zero of the system $F = 0$ by computing at each step the minimal 2-norm solution of $J_F(\bar{x})h = -F(\bar{x})$, where $\bar{x}$ is a suitable point of $\mathbb{R}^n$. Obviously, when the nonlinear system $F = 0$ is ill-conditioned, that is when the condition number of the Jacobian matrix J_F of F evaluated in the constrained region is too elevated, the NFA could end up with unreliable solutions. In order to overcome the numerical instabilities in the computations and avoid the unreliable solutions which can occur in the ill-conditioned case, we present an alternative method that firstly replaces F with a new suitable function $\widehat{F}$ slightly differing from F in the constrained region, that is such that $\|\widehat{F}(x) - F(x)\|_2$ is small on $Q_\varepsilon \cap D$, and well-conditioned at the origin (and also in a neighbourough of it). If this latter condition is not met, that is if the Jacobian matrix $J_{\widehat{F}}(0)$ is ill-conditioned, we prove under a simple additional hypothesis that the original system $F = 0$ subject to constraints has no solution (see Theorem 3.3). Successively our method applies a a modified version of the NFA to the system $\widehat{F} = 0$; the iterations of this modified version are stopped either when the computed deplacement is small enough, as in the classical version, or when the current iteration goes out of the region where $J_{\widehat{F}}$ is well-conditioned. At the end this new algorithm returns either the current iterate $\bar{x}$ if the given constraints are satisfied, or a warning message. This method, that we present as the Root Finding Algorithm (see Algorithm 3.2), turns out to be *backward stable* [8] by construction, since it tries to solve, instead of the original constrained system, the nearby problem $\widehat{F} = 0$ with the same constraints.

In order to define the new function $\widehat{F} : D \to \mathbb{R}^m$, we use the numerical rank of the matrix $(J_F(0) \,|\, F(0))$ as follows. Given a threshold $\delta \geq \varepsilon$ of the same order of magnitude as ε and an integer $k > 1$, we detect the numerical (δ, k)-rank of the matrix $(J_F(0) \,|\, F(0))^t$ by means of the SVD decomposition. Let Π be the permutation matrix whose existence is stated in Theorem 2.8

with $r = \mathrm{rank}_{\delta,k}(J_F(0) \mid F(0))^t$, which satisfies

$$\left(\begin{array}{c} J_F(0)^t \\ F(0)^t \end{array} \right) \Pi = (Q_1 \mid Q_2) \left(\begin{array}{cc} R_{11} & R_{12} \\ 0 & R_{22} \end{array} \right) \tag{3.1}$$

where $Q = (Q_1 \mid Q_2)$ is orthonormal, $Q_1 \in \mathrm{Mat}_{n+1 \times r}(\mathbb{R})$, $Q_2 \in \mathrm{Mat}_{n+1 \times m-r}(\mathbb{R})$, $R_{11} \in \mathrm{Mat}_r(\mathbb{R})$ and $R_{22} \in \mathrm{Mat}_{m-r}(\mathbb{R})$ are upper triangular. Further we have $\sigma_{\min}(R_{11}) \geq (k/q(m,r))\delta$ and $\|R_{22}\|_2 \leq q(m,r)\delta$, where $q(m,r) = \sqrt{r(m-r) + \min(r, m-r)}$. Let $F_1 : D \to \mathbb{R}^r$ and $F_2 : D \to \mathbb{R}^{m-r}$ be the functions made up of the first r and the last $m-r$ components of $\Pi^t F$. The function $\widehat{F} : D \to \mathbb{R}^m$, defined as

$$\widehat{F}(x) \equiv \left(\begin{array}{c} \widehat{F}_1(x) \\ \widehat{F}_2(x) \end{array} \right) = \left(\begin{array}{c} F_1(x) \\ GF_1(x) \end{array} \right) \quad \text{with} \quad G = R_{12}^t R_{11}^{-t} \tag{3.2}$$

is "close" to the function F, as shown in the following theorem.

Theorem 3.1. *Let $\Delta \subseteq Q_\varepsilon \cap D$ be a closed convex set containing the origin and let $\widehat{F}$ be the function defined by (3.2); for each $x \in \Delta$ we have*

$$\|\widehat{F}(x) - \Pi^t F(x)\|_2 \leq q(m,r)\delta + O(\delta^2)$$

Moreover, if there exists $x^* \in \Delta$ such that $F(x^*) = 0$, then

$$\|\widehat{F}(x) - \Pi^t F(x)\|_2 = O(\delta^2)$$

Proof. W.l.o.g. in (3.1) we may assume Π equals to the identity matrix; using (3.1) and an obvious partitioning of $Q = (Q_1 \mid Q_2)$ the matrix $(J_F(0)|F(0))$ can be rewritten as

$$(J_F(0) \mid F(0)) = \left(\begin{array}{cc} J_{F_1}(0) & F_1(0) \\ J_{F_2}(0) & F_2(0) \end{array} \right) = \left(\begin{array}{c} R_{11}^t Q_1^t \\ R_{12}^t Q_1^t + R_{22}^t Q_2^t \end{array} \right) \tag{3.3}$$

We denote by $d(x) = (\widehat{F}_2 - F_2)(x) = (GF_1 - F_2)(x)$ and so from (3.2) we have that $\|\widehat{F}(x) - F(x)\|_2 = \|\widehat{F}_2(x) - F_2(x)\|_2 = \|d(x)\|_2$. Further from (3.3)

$$\begin{aligned} (J_d(0) \mid d(0)) &= (GJ_{F_1}(0) - J_{F_2}(0) \mid GF_1(0) - F_2(0)) \\ &= G(J_{F_1}(0) \mid F_1(0)) - (J_{F_2}(0) \mid F_2(0)) \\ &= GR_{11}^t Q_1^t - R_{12}^t Q_1^t - R_{22}^t Q_2^t = -R_{22}^t Q_2^t \end{aligned}$$

therefore $\|J_d(0)\|_2 \leq q(m,r)\delta$ and $\|d(0))\|_2 \leq q(m,r)\delta$.

11

We consider the Taylor expansion of $d(x)$ at the origin; since $d \in \mathcal{C}^2(\Delta)$ and $\Delta \subseteq Q_\varepsilon \cap D$ is a closed and convex set, we have $d(x) = d(0) + J_d(0)x + O(\varepsilon^2)$, for each $x \in \Delta$. Since $\|x\|_2 \leq \sqrt{n}\varepsilon$ and $\delta \geq \varepsilon$ we have

$$
\begin{aligned}
\|d(x)\|_2 &\leq \|d(0)\|_2 + \|J_d(0)\|_2\|x\|_2 + O(\varepsilon^2) \\
&\leq \|d(0)\|_2 + q(m,r)\sqrt{n}\delta\varepsilon + O(\varepsilon^2) \\
&\leq \|d(0)\|_2 + q(m,r)\sqrt{n}\delta^2 + O(\delta^2) = \|d(0)\|_2 + O(\delta^2)
\end{aligned}
$$

From $\|d(0))\|_2 \leq q(m,r)\delta$ the first inequality follows.

Now, let $x^* \in \Delta$ such that $F(x^*) = 0$; evaluating the previous Taylor expansion of $d(x)$ at x^* we get $0 = d(0) + J_d(0)x^* + O(\varepsilon^2)$, and so $d(0) = -J_d(0)x^* + O(\varepsilon^2)$. Since $\|x^*\|_2 \leq \sqrt{n}\varepsilon$, it follows that

$$
\|d(0)\|_2 \leq q(m,r)\sqrt{n}\delta\varepsilon + O(\varepsilon^2) \leq q(m,r)\sqrt{n}\delta^2 + O(\delta^2) = O(\delta^2)
$$

and this concludes the proof. $\qquad\square$

From Theorem 3.1, the partition given in (3.1) can be used in the Root Finding Algorithm to derive the new function $\widehat{F}$.

Since, by construction, the last $m - r$ equations of the nonlinear system $\widehat{F} = 0$ linearly depend on the first r equations, in order to approximate the solution of $\widehat{F} = 0$ we only have to consider the nonlinear system $F_1 = 0$.

From formula (2.2), since $(J_{F_1}(0) \mid F_1(0)) = R_{11}^t Q_1^t$ it follows that

$$
\sigma_r \left((J_{F_1}(0) \mid F_1(0)) \right) = \sigma_{\min}(R_{11}) > \frac{\sigma_r \left((J_F(0) \mid F(0)) \right)}{q(m,r)}
$$

and so $(J_{F_1}(0) \mid F_1(0))$ is well-conditioned if the r-th singular value of the matrix $(J_F(0) \mid F(0))$ is large enough. Nevertheless this fact does not guarantee that the matrix $J_{F_1}(0)$ is well conditioned too. For this reason, the Root Finding Algorithm, after the construction the function F_1, checks whether the numerical rank of $J_{F_1}(0)$ is equal to r. If $\text{rank}_{\delta,k}(J_{F_1}(0)) < r$ then $J_{F_1}(0)$ is ill-conditioned (see Remark 2.5) and, in order to avoid numerical instabilities in the computations, no iterations are performed. Otherwise, that is if $\text{rank}_{\delta,k}(J_{F_1}(0)) = r$ the matrix $J_{F_1}(0)$ is well-conditioned; in this case, starting from the zero vector, the Root Finding Algorithm processes the system $F_1 = 0$ by mean the NFA with an additional check, at each step, on the conditioning of the Jacobian matrix, in order to deal with well conditioned linear systems. When the loop stops, the algorithm returns the final iterate if it satisfies the constraints or, otherwise, a warning message.

Algorithm 3.2. (The Root Finding Algorithm – RFA)
Let $D \subseteq \mathbb{R}^n$ be a set containing the origin and $F : D \to \mathbb{R}^m$ with $m < n$ be a nonlinear function of class $\mathcal{C}^2(D)$; let $\omega \ll 1$ be a fixed threshold, $\varepsilon > 0$ and $Q_\varepsilon = \{x \in \mathbb{R}^n \ : \ \|x\|_\infty \leq \varepsilon\}$; let $\delta \geq \varepsilon$ of the same order of magnitude as ε and $k > 1$. Consider the following sequence of steps.

RF1 Compute the numerical (δ, k)-rank of $(J_F(0)\,|\,F(0))^t$ and a permutation matrix $\Pi \in \mathrm{Mat}_m(\mathbb{R})$ as given by Theorem 2.8 w.r.t. the numerical rank$r = \mathrm{rank}_{\delta,k}\left((J_F(0)\,|\,F(0))^t\right)$. Partition Π as $(\Pi_1\,|\,\Pi_2)$ with $\Pi_1 \in \mathrm{Mat}_{m\times r}(\mathbb{R})$; let $F_1 = \Pi_1^t F$.

RF2 Let $\bar{x} = 0$, $h = (1,\ldots,1)^t$.
While $(\|h\|_2 > \omega)$ and $(\mathrm{rank}_{\delta,k}(J_{F_1}(\bar{x})) = r)$
  • Compute the minimal 2-norm solution h of $J_{F_1}(\bar{x})h = -F_1(\bar{x})$.
  • Let $\bar{x} = \bar{x} + h$.

RF3 If $\bar{x} \in Q_\varepsilon \cap D$ return $\bar{x}$; otherwise return $\bar{x} =$NULL.

Step **RF1** requires the determination of the permutation matrix Π; the existence of Π is guaranteed by Theorem 2.8, its computation can be performed using a Rank-Revealing QR Factorization [14]. We refer to [3], [4] for a detailed description of algorithms for computing such a decomposition.

Note that, if $J_{F_1}(0)$ is well-conditioned then there exists a closed neighbourough Δ_0 of the origin contained in $Q_\varepsilon \cap D$ such that $J_{F_1}(x)$ is well-conditioned for each $x \in \Delta_0$. In fact, since $F_1 \in \mathcal{C}^2(D)$, then $J_{F_1}(x)$ is a Lipschitz function (with constant γ) on Δ_0 and so, from Proposition 2.7, for each $x \in \Delta_0$ we have that $|\sigma_r(J_{F_1}(x)) - \sigma_r(J_{F_1}(0))| \leq \|J_{F_1}(x) - J_{F_1}(0)\|_2 \leq \gamma\|x\|_2$ which implies $\sigma_r(J_{F_1}(x)) \geq \sigma_r(J_{F_1}(0)) - \gamma\|x\|_2$. Since $\sigma_r(J_{F_1}(0)) > k\delta$, if $\|x\|_2$ is small enough, that is if x belongs to a suitable neighbourough Δ_0 of the origin, then $J_{F_1}(x)$ has full numerical (δ, k)-rank on Δ_0 and so the RFA executes some iterations at step **RF2** for approximating the solution of $F_1 = 0$.

We analyze the output of the RFA; if the output $\bar{x} \neq$ NULL then the peculiarities of $\bar{x}$ give us different information on the underdetermined nonlinear system $F = 0$. If $\bar{x} = 0$ is returned, then two possible cases can occur: either a single iteration or no iteration of the loop of step **RF2** has been performed. In the former case h is the zero vector: it means that $F_1(\bar{x}) = 0$ and so $\widehat{F}(\bar{x}) = 0$, that is the exact solution of $\widehat{F} = 0$ has been found. From Theorem 3.1 it follows that $\|F(\bar{x})\|_2 = \|F(\bar{x}) - \widehat{F}(\bar{x})\|_2 < q(m,r)\delta$. In the latter case no iteration has been executed since $\mathrm{rank}_{\delta,k}(J_{F_1}(0)) < r$, though

$\mathrm{rank}_{\delta,k}(J_F(0) \mid F(0)) = r$; in this situation $h = (1 \ldots 1)^t$ and, if some simple hypotheses on a closed convex set $\Delta \subseteq Q_\varepsilon \cap D$ are satisfied, Theorem 3.3 shows that there are no solutions of $F = 0$ in Δ.

Theorem 3.3. *Let $\Delta \subseteq Q_\varepsilon \cap D$ be a closed convex set containing the origin and let γ be the Lipschitz constant of J_{F_1} on Δ. Suppose that the matrix $(J_{F_1}(0) \mid F_1(0))$ has full numerical (δ, k)-rank r with $k > 1 + (\sqrt{n} + n\gamma)\delta$. If $\mathrm{rank}_{\delta,k}(J_{F_1}(0)) < r$ then there are no solutions of $F = 0$ in Δ.*

Proof. Note that, since $F \in \mathcal{C}^2(\Delta)$ and Δ is a closed convex set, J_F is a Lipschitz function on Δ. Let $r_1 = \mathrm{rank}_{\delta,k}(J_{F_1}(0))$ and $A \in \mathrm{Mat}_{r \times n}(\mathbb{R})$ be such that $\mathrm{rank}(A) = r_1$ and $\|J_{F_1}(0) - A\|_2 < \delta$: the existence of such a matrix A follows from Lemma 2.6. Since A has exact rank r_1 there exist a permutation matrix $\Pi \in \mathrm{Mat}_r(\mathbb{R})$ and a matrix $W \in \mathrm{Mat}_{(r-r_1) \times r_1}(\mathbb{R})$ such that

$$\Pi A = \begin{pmatrix} A_1 \\ W A_1 \end{pmatrix} \begin{matrix} \}r_1 \\ \}r - r_1 \end{matrix} \quad \text{and} \quad \mathrm{rank}(A_1) = r_1$$

We can express the matrix $\Pi J_{F_1}(0)$ as follows

$$\Pi J_{F_1}(0) = \begin{pmatrix} A_1 + E_{11} \\ W A_1 + E_{21} \end{pmatrix} \quad \text{with} \quad E = \begin{pmatrix} E_{11} \\ E_{21} \end{pmatrix} \quad \text{and} \quad \|E\|_2 < \delta$$

Suppose for a contradiction that there exists $x^* \in \Delta$ such that $F(x^*) = 0$, and so $F_1(x^*) = 0$. From Taylor's theorem applied to $F_1(x)$ we know that there exists a point z on the line connecting 0 to x^*, such that $0 = F_1(x^*) = F_1(0) + J_{F_1}(z)x^*$, and so

$$0 = \Pi F_1(0) + \Pi J_{F_1}(z)x^* = \Pi F_1(0) + \Pi J_{F_1}(0)x^* + \Pi(J_{F_1}(z) - J_{F_1}(0))x^*$$

Using an obvious partitioning $(F_{11}, F_{12})^t$ and $(E_{12}, E_{22})^t$ of the matrices $\Pi F_1(0)$ and $\Pi(J_{F_1}(z) - J_{F_1}(0))$ we obtain

$$0 = \begin{pmatrix} F_{11} \\ F_{12} \end{pmatrix} + \begin{pmatrix} A_1 + E_{11} \\ W A_1 + E_{21} \end{pmatrix} x^* + \begin{pmatrix} E_{12} \\ E_{22} \end{pmatrix} x^*$$

and therefore

$$\begin{cases} A_1 x^* &= -F_{11} - (E_{11} + E_{12})x^* \\ F_{12} &= W(F_{11} + (E_{11} + E_{12})x^*) - (E_{21} + E_{22})x^* \end{cases}$$

It follows that $\Pi\left(J_{F_1}(0) \mid F_1(0)\right)$ is equal to the matrix

$$\begin{pmatrix} A_1 + E_{11} & F_{11} \\ W A_1 + E_{21} & W\left(F_{11} + (E_{11} + E_{12})x^*\right) - (E_{21} + E_{22})x^* \end{pmatrix}$$

Let $B \in \mathrm{Mat}_{r \times n}(\mathbb{R})$ such that

$$\Pi B = \begin{pmatrix} A_1 & F_{11} + (E_{11} + E_{12})x^* \\ W A_1 & W\left(F_{11} + (E_{11} + E_{12})x^*\right) \end{pmatrix}$$

since $\|\cdot\|_2$ is invariant under permutation and $x^*, z \in \Delta$ we have

$$\begin{aligned} \|B - (J_{F_1}(0) \mid F_1(0))\|_2 &= \|(-E \mid Ex^* + \Pi(J_{F_1}(z) - J_{F_1}(0))x^*)\|_2 \\ &\leq \|E\|_2 (1 + \|x^*\|_2) + \|J_{F_1}(z) - J_{F_1}(0)\|_2 \|x^*\|_2 \\ &\leq \delta(1 + \sqrt{n}\varepsilon) + \gamma n \varepsilon^2 \leq \delta + (\sqrt{n} + n\gamma)\delta^2 \end{aligned}$$

From $\mathrm{rank}(A_1) = r_1$ we have that $\mathrm{rank}(B) = r_1$, and so $\sigma_r(B) = 0$. From Proposition 2.7 it follows that $\sigma_r\left(J_{F_1}(0) \mid F_1(0)\right) \leq \delta + (\sqrt{n} + n\gamma)\delta^2$. Since from the hypothesis $\sigma_r(J_{F_1}(0) \mid F_1(0)) > k\delta$ and $k > 1 + (\sqrt{n} + n\gamma)\delta$ we obtain a contradiction. $\qquad\square$

If the RFA returns a non zero vector $\bar{x} \in Q_\varepsilon \cap D$ then $F(\bar{x})$ assumes small values; in particular, $\|F(\bar{x})\|_2$ is bounded by a function depending on the data tolerance as the following theorem shows.

Theorem 3.4. *Let $\bar{x}$ be the output of the RFA applied to $F \in \mathcal{C}^2(D)$ and H be the closed convex hull of the sequence of points computed by the RFA. If $H \subseteq Q_\varepsilon \cap D$ then*

$$\|\widehat{F}(\bar{x})\|_2 = O(\delta^2) \quad and \quad \|F(\bar{x})\|_2 \leq q(m, r)\delta + O(\delta^2)$$

Further, let Δ be a closed convex set such that $H \subseteq \Delta \subseteq Q_\varepsilon \cap D$; if there exists $x^* \in \Delta$ such that $F(x^*) = 0$ then $\|F(\bar{x})\|_2 = O(\delta^2)$.

Proof. Let h be the displacement computed in the last iteration of the RFA, that is let h satisfies $J_{F_1}(\bar{x} - h)h = -F_1(\bar{x} - h)$ and consequently $J_{\widehat{F}}(\bar{x} - h)h = -\widehat{F}(\bar{x} - h)$. Since $\bar{x} - h \in H$, $\widehat{F} \in \mathcal{C}^2(H)$ and H is a closed convex set, we express $\widehat{F}(\bar{x})$ using the Taylor expansion of $\widehat{F}$ centered at $\bar{x} - h$ we have $\widehat{F}(\bar{x}) = \widehat{F}(\bar{x} - h) + J_{\widehat{F}}(\bar{x} - h)h + O(\|h\|^2) = O(\|h\|^2)$ and, from $\|h\|_2 \leq \sqrt{n}\varepsilon \leq \sqrt{n}\delta$, it follows that $\|\widehat{F}(\bar{x})\|_2 = O(\delta^2)$.

Since $\|F(\bar{x})\|_2 = \|\Pi F(\bar{x})\|_2 \leq \|\widehat{F}(\bar{x}) - \Pi F(\bar{x})\|_2 + \|\widehat{F}(\bar{x})\|_2$, from Theorem 3.1 we have $\|F(\bar{x})\|_2 \leq q(m, r)\delta + O(\delta^2)$ and, if there exists $x^* \in \Delta$ such that $F(x^*) = 0$ then $\|F(\bar{x})\|_2 = O(\delta^2)$. $\qquad\square$

15

4 The Low-degree Polynomial Algorithm

Given a set $\mathbb{X}^\varepsilon$ of s distinct empirical points of $\mathbb{R}^n$ in this paper we address the problem to determine a polynomial of $P = \mathbb{R}[x_1, \ldots, x_n]$ whose degree is bounded by the minimal degree of the elements of the vanishing ideal $\mathcal{I}(\mathbb{X})$ and whose affine variety exactly contains an admissible perturbation $\mathbb{X}^*$ of $\mathbb{X}^\varepsilon$.

In this section we present an algorithm that computes an admissible perturbation $\overline{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$ and a polynomial f, whose degree is bounded by the minimal degree of the elements of $\mathcal{I}(\mathbb{X})$; since f has low degree and, by construction, assumes small values at $\overline{\mathbb{X}}$, it is a good candidate to be the solution to the addressed problem. In fact, if some simple additional hypotheses are satisfied, f turns out to be the polynomial we are looking for, as in Theorem 4.3 we prove that there exists an admissible perturbation $\mathbb{X}^*$, slightly differing from $\overline{\mathbb{X}}$, contained in the affine variety $\mathcal{Z}(f)$.

As already pointed out in the Introduction a straightforward application of the Buchberger-Möller (BM) algorithm to empirical data provides unreliable results, because the problem of deciding whether a polynomial vanishes at the input points is very sensitive to perturbations. Nevertheless the strategy of the BM algorithm may be adapted to limited precision data; for instance the NBM and SOI algorithms ([2], [10]) employ such an approach to compute polynomials which almost vanish at the input points. Also in this context we present a modified version of the BM algorithm.

The core of the BM algorithm is the stepwise construction of the monomial basis $\mathcal{O}$ of the quotient ring $P/\mathcal{I}(\mathbb{X})$ viewed as an $\mathbb{R}$-vector space. Initially $\mathcal{O}$ comprises just the power product 1; at the generic iteration $\mathcal{O} = \{t_1, \ldots, t_\nu\}$ and, fixing a term ordering τ, the smallest $t >_\tau t_i$, $i = 1 \ldots \nu$, is considered. If the evaluation vector $t(\mathbb{X})$ is linearly dependent on the columns of the matrix $M_{\mathcal{O}}(\mathbb{X})$, that is if

$$\rho(\mathbb{X}) = 0 \tag{4.1}$$

where $\rho(\mathbb{X}) = t(\mathbb{X}) - M_{\mathcal{O}}(\mathbb{X})\alpha(\mathbb{X})$ is the residual of the least squares problem $M_{\mathcal{O}}(\mathbb{X})\alpha(\mathbb{X}) = t(\mathbb{X})$, then the polynomial $t - \sum_{i=1}^{\nu} \alpha_i(\mathbb{X})t_i$ is formed and added to a basis, otherwise t is added to $\mathcal{O}$.

We adapt the generic iteration of the BM algorithm to our case by replacing check (4.1) by

$$\rho(\mathbb{X}^*) = 0 \text{ for some admissible perturbation } \mathbb{X}^* \text{ of } \mathbb{X}^\varepsilon \tag{4.2}$$

If problem (4.2) admits solution then the polynomial $f^* = t - \sum_{i=1}^{\nu} \alpha_i(\mathbb{X}^*)t_i$ is formed; f^* and $\mathbb{X}^*$ are a solution we were looking for (by construction f^* vanishes at $\mathbb{X}^*$). Otherwise t is added to $\mathcal{O}$.

Analogously to the strategy used in [2], in order to make problem (4.2) effectively solvable we express any admissible perturbation of $\mathbb{X}^\varepsilon$ as a function of sn **error variables** $\mathbf{e} = (e_{11}, \ldots, e_{s1}, e_{12}, \ldots, e_{s2}, \ldots, e_{1n}, \ldots, e_{sn})$, where each variable e_{kj} represents the perturbation in the j-th coordinate of the specified value $p_k \in \mathbb{X}$. Specifically, a generic admissible perturbation $\widetilde{\mathbb{X}}$ of $\mathbb{X}^\varepsilon$ can be expressed as $\widetilde{\mathbb{X}} = \mathbb{X}(\mathbf{e}) = \{p_1(\mathbf{e}_1), \ldots, p_s(\mathbf{e}_s)\}$ where for each $k = 1 \ldots s$

$$\mathbf{e}_k = (e_{k1}, \ldots, e_{kn})$$
$$p_k(\mathbf{e}_k) = (p_{k1} + e_{k1},\ p_{k2} + e_{k2}, \ldots p_{kn} + e_{kn}) \quad \text{and} \quad \|\mathbf{e}_k\|_\infty \le \varepsilon$$

Using this formalization, $\widetilde{\mathbb{X}} = \mathbb{X}(\mathbf{e})$ is a function in the variables $\mathbf{e}$ with admissible domain $Q_\varepsilon = \{\mathbf{e} \in \mathbb{R}^{sn} : \|\mathbf{e}\|_\infty \le \varepsilon\}$. It follows that $\alpha(\widetilde{\mathbb{X}})$, $\rho(\widetilde{\mathbb{X}})$, and $M_\mathcal{O}(\widetilde{\mathbb{X}})$ are functions of $\mathbf{e}$, simply denoted by $\alpha(\mathbf{e})$, $\rho(\mathbf{e})$ and $M_\mathcal{O}(\mathbf{e})$; in particular $\rho : \mathbb{R}^{sn} \to \mathbb{R}^s$ is defined by

$$\rho(\mathbf{e}) = t(\mathbf{e}) - M_\mathcal{O}(\mathbf{e})\alpha(\mathbf{e}) \tag{4.3}$$

where $\alpha(\mathbf{e}) = M_\mathcal{O}(\mathbf{e})^\dagger t(\mathbf{e})$. It follows that problem (4.2) is equivalent to determine whether

$$\rho(\mathbf{e}) = 0 \text{ for some } \mathbf{e} \in Q_\varepsilon \cap D_\mathcal{O} \tag{4.4}$$

where $D_\mathcal{O}$ is the domain of ρ, that is $D_\mathcal{O} = \{\mathbf{e} \in \mathbb{R}^{sn} : M_\mathcal{O}(\mathbf{e}) \text{ is full rank}\}$.

In [10] it is proved that the problem (4.4) has no solution if

$$|\rho(0)| > \varepsilon |I - M_\mathcal{O}(0)M_\mathcal{O}(0)^\dagger| \sum_{k=1}^{n} |\partial_k t(0) - M_{\partial_k \mathcal{O}}(0)\alpha(0)| + O(\varepsilon^2) \tag{4.5}$$

where the absolute value of a matrix is the matrix consisting of the absolute values of its elements and the lower bound means that the relation holds componentwise. In our algorithm we employ such a result to study problem (4.4): we solve the underdetermined nonlinear system (4.4) only if condition (4.5) is not satisfied. Unfortunately, finding an exact solution of problem (4.4) can be computationally very hard: in fact the analytic expression of $\rho(\mathbf{e})$ can be very hard to obtain and even when an explicit expression of $\rho(\mathbf{e})$ is available,

an exact solution $\mathbf{e}^*$ of $\rho(\mathbf{e}) = 0$ can be computed only in very few cases (specifically when n and s are small), while in general only an approximation of $\mathbf{e}^*$ can be effectively determined. In our algorithm we approximate the solution of problem (4.4) using the results of Section 3; in particular, we design the The Low-degree Polynomial Algorithm following the frame of the BM algorithm and solving the nonlinear systems (4.4) using the RFA.

Algorithm 4.1. (The Low-degree Polynomial Algorithm - LPA)
Let $\mathbb{X}^\varepsilon = \{p_1^\varepsilon, \ldots, p_s^\varepsilon\}$ be a finite set of distinct empirical points of $\mathbb{R}^n$, $k > 1$, $\delta \geq \varepsilon$ of the same magnitude as ε and τ be a term ordering.
Consider the following sequence of steps.

LP1 Start with $t = 1$ and the lists $\mathcal{O} = [1]$ and $L = [\,]$.

LP2 Add to L those elements of $\{x_1 t, \ldots, x_n t\}$ which are not multiples of elements of L. Let $t = \min_\tau(L)$ and delete it from L.

LP3 Choose $\bar{\mathbf{e}} = 0$ and consider the residual function ρ defined in (4.3).

LP4 If $\rho(0)$ satisfies condition (4.5) then
> • add t to the list $\mathcal{O}$ and continue with step **LP2**;
> else
> • apply the RFA to ρ on $Q_\varepsilon \cap D_\mathcal{O}$ with thresholds k and δ.

LP5 If the RFA returns NULL or $\bar{\mathbf{e}} = 0$ and $\|h\| \neq 0$ then
> • add t to the list $\mathcal{O}$ and continue with step **LP2**;
> else
> • compute $f = t - \sum_{t_i \in \mathcal{O}} \alpha_i(\bar{\mathbf{e}}) t_i$; return $\bar{\mathbf{e}}$, f, $\mathcal{O}$ and stop.

We make a few observations on the LPA. At each iteration the domain $D_\mathcal{O}$ of the residual function ρ contains the origin. In fact since the residuals computed at the previous iterations do not vanish at zero the matrix $M_\mathcal{O}(0)$ is full rank. Further, at step **LP4** each iteration of the RFA requires the evaluation of ρ and J_ρ at a given point $\mathbf{e}$. The value of $\rho(\mathbf{e})$ is simply obtained by solving the numerical least squares problem $M_\mathcal{O}(\mathbf{e})\alpha(\mathbf{e}) = t(\mathbf{e})$ and computing its residual. Proposition 6.1 illustrates how to compute each column of the Jacobian matrix $J_\rho(\mathbf{e})$ without passing through the explicit expression of the function ρ.

Theorem 4.2. *The Low-degree Polynomial Algorithm stops after finitely many steps and returns $\bar{\mathbf{e}}$, f, $\mathcal{O}$ such that $\bar{\mathbf{e}} \in Q_\varepsilon \cap D_\mathcal{O}$.*

18

Let ρ be the residual function computed during the last iteration of the LPA and H be the closed convex hull of the sequence of points computed by the RFA applied to ρ. If $H \subseteq Q_\varepsilon \cap D_\mathcal{O}$ then

$$\|f(\mathbb{X}(\bar{\mathbf{e}}))\|_2 \leq q(s,r)\delta + O(\delta^2)$$

where $r = \mathrm{rank}_{\delta,k}(J_\rho(0) \mid \rho(0))$ and $q(s,r) = \sqrt{r(s-r) + \min(r, s-r)}$.

Further, let Δ be a closed convex set such that $H \subseteq \Delta \subseteq Q_\varepsilon \cap D_\mathcal{O}$; if there exists $\mathbf{e}^* \in \Delta$ such that $\rho(\mathbf{e}^*) = 0$ then $\|f(\mathbb{X}(\bar{\mathbf{e}}))\|_2 = O(\delta^2)$

Proof. At each iteration either the algorithm computes the polynomial f and thus it stops, or a term t is added to $\mathcal{O}$. We observe that this latter instruction can be executed at most $s - 1$ times; in fact when $\mathcal{O}$ contains s terms, that is when $M_\mathcal{O}(0)$ becomes a square matrix, the residual vector $\rho(0)$ is zero, and consequently $\bar{\mathbf{e}} = 0$, $\rho(\bar{\mathbf{e}}) = 0$, and a polynomial f is computed. From the check of step **LP5** obviously follows that $\bar{\mathbf{e}} \in Q_\varepsilon \cap D_\mathcal{O}$.

The residual function $\rho \in \mathcal{C}^\infty(D_\mathcal{O})$ since each of its components is a rational function defined in $D_\mathcal{O}$. The thesis follows from Theorem 3.4 applied to ρ and the equality $f(\mathbb{X}(\bar{\mathbf{e}})) = \rho(\bar{\mathbf{e}})$. $\qquad\square$

In the following theorem, under additional hypotheses on the polynomial f, output of the LPA, we prove that there exists an admissible perturbation $\mathbb{X}(\mathbf{e}^*)$ of $\mathbb{X}^\varepsilon$ contained in $\mathcal{Z}(f)$; further, we provide an estimation of the distance between $\mathbb{X}(\bar{\mathbf{e}})$ and $\mathbb{X}(\mathbf{e}^*)$ which is usually much smaller than ε.

Theorem 4.3. *Let $(\bar{\mathbf{e}}, f, \mathcal{O})$ be the output of the LPA applied to $\mathbb{X}^\varepsilon$. For each $i = 1 \ldots s$ let $f_i : \mathbb{R}^n \to \mathbb{R}$ be the function defined by $f_i(\mathbf{e}_i) = f(p_i(\mathbf{e}_i))$; let r_i and R_i be real positive numbers such that*

$$r_i = \varepsilon - \|\bar{\mathbf{e}}_i\|_\infty \quad and \quad R_i < \min\left\{r_i, \frac{\|J_{f_i}(\bar{\mathbf{e}}_i)\|_2}{\gamma_i}\right\}$$

where $\gamma_i > 0$ is a Lipschitz constant of J_{f_i} in $B(\bar{\mathbf{e}}_i, r_i)$. Let μ_i be an upper bound of $\|J_{f_i}^\dagger\|_2$ in $B(\bar{\mathbf{e}}_i, R_i)$. If $\|J_{f_i}(\bar{\mathbf{e}}_i)\|_2 > 0$ and

$$|f_i(\bar{\mathbf{e}}_i)| < \frac{R_i}{\mu_i(2 + \gamma_i R_i \mu_i)} \equiv \chi_i \qquad \forall i = 1 \ldots s \qquad (4.6)$$

then there exists $\mathbf{e}^*$ such that the set $\mathbb{X}(\mathbf{e}^*)$ is an admissible perturbation of $\mathbb{X}^\varepsilon$, $f(\mathbb{X}(\mathbf{e}^*)) = 0$, and $\|\mathbf{e}^* - \bar{\mathbf{e}}\|_\infty < \max R_i$.

Proof. Let $\bar{\alpha}_j = \alpha_j(\bar{\mathbf{e}})$ and $f = t - \sum_{t_j \in \mathcal{O}} \bar{\alpha}_j t_j$. We observe that $f_i(\mathbf{e}_i) = t(p_i(\mathbf{e}_i)) + \sum_{t_j \in \mathcal{O}} \bar{\alpha}_j t_j(p_i(\mathbf{e}_i))$ is a polynomial function of $\mathbf{e}_i$ and its Jacobian $J_{f_i}(\mathbf{e}_i) \in \mathrm{Mat}_{1 \times n}(\mathbb{R})$ is a Lipschitz function in $\bar{B}(\bar{\mathbf{e}}_i, r_i)$. We prove that $\|J_{f_i}(\mathbf{e}_i)\|_2 > 0$ in $\bar{B}(\bar{\mathbf{e}}_i, R_i)$. For each $\mathbf{e}_i \in \bar{B}(\bar{\mathbf{e}}_i, R_i)$ we have

$$| \|J_{f_i}(\mathbf{e}_i)\|_2 - \|J_{f_i}(\bar{\mathbf{e}}_i)\|_2 | \leq \|J_{f_i}(\mathbf{e}_i) - J_{f_i}(\bar{\mathbf{e}}_i)\|_2 \leq \gamma_i \|\mathbf{e}_i - \bar{\mathbf{e}}_i\|_2 \leq \gamma_i R_i$$

and so $\|J_{f_i}(\mathbf{e}_i)\|_2 \geq \|J_{f_i}(\bar{\mathbf{e}}_i)\|_2 - \gamma_i R_i > 0$. It follows that $J_{f_i}(\mathbf{e}_i)$ has constant rank 1 in $\bar{B}(\bar{\mathbf{e}}_i, R_i)$; further, $\|J_{f_i}^\dagger(\mathbf{e}_i)\|_2$ is upper bounded in $\bar{B}(\bar{\mathbf{e}}_i, R_i)$ since

$$\|J_{f_i}^\dagger(\mathbf{e}_i)\|_2 = \frac{\|J_{f_i}^t(\mathbf{e}_i)\|_2}{\|J_{f_i}(\mathbf{e}_i)\|_2^2} = \frac{1}{\|J_{f_i}(\mathbf{e}_i)\|_2} \leq \frac{1}{\min_{\bar{B}(\bar{\mathbf{e}}_i, R_i)} \|J_{f_i}(\mathbf{e}_i)\|_2}$$

and $\min_{\bar{B}(\bar{\mathbf{e}}_i, R_i)} \|J_{f_i}(\mathbf{e}_i)\|_2 > 0$.

Now we apply Kantorovich theorem to f_i on $B(\bar{\mathbf{e}}_i, R_i)$ with $x_0 = \bar{\mathbf{e}}_i$, $p = 1$, $\eta = R_i$ and so $\Omega_\eta = \{\bar{\mathbf{e}}_i\}$; since $\chi_i = \frac{R_i}{\mu_i(2 + \gamma_i R_i \mu_i)}$ satisfies $\frac{\gamma_i \mu_i^2 \chi_i}{2} < 1$ and $\frac{2\mu_i \chi_i}{(2 - \gamma_i \mu_i^2 \chi_i)} < R_i$, and (4.6) holds we conclude there exists $\mathbf{e}_i^* = (e_{i1}^* \ldots e_{in}^*)$ in $B(\bar{\mathbf{e}}_i, R_i)$ such that $\bar{f}_i(\mathbf{e}_i^*) = 0$. Let $\mathbf{e}^* = (e_{11}^* \ldots e_{s1}^*, \ldots, e_{1n}^* \ldots e_{sn}^*)$; since $f(\mathbb{X}(\mathbf{e}^*)) = (f_1(\mathbf{e}_1^*), \ldots, f_s(\mathbf{e}_s^*))^t$ we have $f(\mathbb{X}(\mathbf{e}^*)) = 0$. Furthermore, since the i-th element of $\mathbb{X}$ is $p_i = p_i(0)$ and

$$\|p_i(\mathbf{e}_i^*) - p_i(0)\|_\infty = \|\mathbf{e}_i^*\|_\infty \leq \|\mathbf{e}_i^* - \bar{\mathbf{e}}_i\|_2 + \|\bar{\mathbf{e}}_i\|_\infty \leq R_i + \|\bar{\mathbf{e}}_i\|_\infty < \varepsilon$$

we have that $\mathbb{X}(\mathbf{e}^*)$ is an admissible perturbation of $\mathbb{X}^\varepsilon$. Finally, we have that $\|\mathbf{e}^* - \bar{\mathbf{e}}\|_\infty = \max_i \|\mathbf{e}_i^* - \bar{\mathbf{e}}_i\|_\infty \leq \max_i \|\mathbf{e}_i^* - \bar{\mathbf{e}}_i\|_2 < \max_i R_i < \varepsilon$ which concludes the proof. $\square$

5 Numerical examples

In this section we present some numerical examples to show the effectiveness of the LPA. We implemented the LPA and the RFA using the C++ language, the CoCoALib [6], and some routines of GSL - GNU Scientific Library [12]; all computations have been performed on an Intel Core 2 Duo processor (at 1.86 GHz). In all the examples the degree lexicographic term ordering with $y < x$ is understood; furthermose the coordinates of the points and the coefficients of the polynomials are sometimes displayed as truncated decimals.

In Example 5.1 the LPA is applied to the set of points of Example 1.1 created by perturbing by less than 0.1 the coordinates of 10 points located

on the parabola $g = 0$ where $g = y^2 - x - 2y + 2$. In contrast to the exact approach that provides a minimal degree polynomial of degree 4, the LPA detects the geometrical configuration given by a parabola.

Example 5.1. Let $\mathbb{X} \subset \mathbb{R}^2$ be the set of points given in Example 1.1, let $\varepsilon = 0.1$, $\delta = 2\varepsilon$ and $k = 2$. The LPA applied to $\mathbb{X}^\varepsilon$ performs three iterations. At the first two iterations the terms $t = y$ and $t = x$ are analyzed: since in both cases condition (4.5) is satisfied the set $\mathcal{O} = \{1, y, x\}$ is constructed. During the third iteration the term $t = y^2$ is considered: at step **LP4** the RFA applied to the current residual function ρ returns $\bar{\mathbf{e}} \in Q_\varepsilon \cap D_\mathcal{O}$, therefore at step **LP5** a polynomial f is computed (see Figure 1), and the algorithm stops with the following output:

$$
\begin{aligned}
\bar{\mathbf{e}} \;=\; & 10^{-1}(0.474, -0.211, -0.146, -0.100, 0.075, -0.105, -0.091, -0.017, \\
& 0.066, 0.054, 0.004, -0.215, 0.285, -0.202, 0.211, 0.378, -0.332, \\
& -0.065, -0.369, 0.305) \\
f \;=\; & y^2 - 0.9751012065x - 2.0049270587y + 1.9775224038 \\
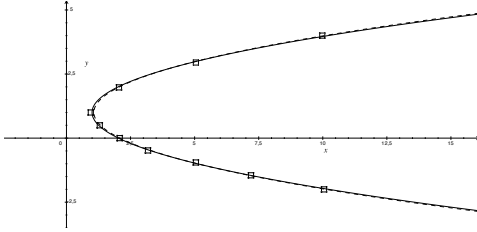\mathcal{O} \;=\; & \{1, y, x\}
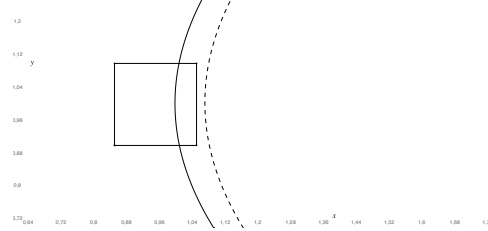\end{aligned}
$$

Figure 1: Variety $\mathcal{Z}(f)$

Figure 2: The varieties $\mathcal{Z}(f)$ and $\mathcal{Z}(h)$: zoom in around the point (0.95, 1)

The set $\mathbb{X}(\bar{\mathbf{e}})$ is an admissible perturbation of $\mathbb{X}^\varepsilon$; we can consider the polynomial f almost vanishing at $\mathbb{X}(\bar{\mathbf{e}})$ since $\|f(\mathbb{X}(\bar{\mathbf{e}}))\|_2/\|f\| < 2 \cdot 10^{-5}$. We apply Theorem 4.3 to this example and report the values of R_i, χ_i and $|f_i(\bar{\mathbf{e}}_i)|$ in the following table. Since for each i the inequality (4.6) is satisfied, we conclude that there exists $\mathbf{e}^*$ such that $\mathbb{X}(\mathbf{e}^*)$ is an admissible perturbation of $\mathbb{X}^\varepsilon$, $f(\mathbb{X}(\mathbf{e}^*)) = 0$ and $\|\mathbf{e}^* - \bar{\mathbf{e}}\|_\infty < 0.0935$.

	i=1	i=2	i=3	i=4	i=5
R_i	0.0526	0.0785	0.0715	0.0798	0.0789
χ_i	0.0243	0.0354	0.0325	0.0360	0.0356
$|f_i(\bar{\mathbf{e}}_i)|$	$3.3 \cdot 10^{-7}$	$3.87 \cdot 10^{-6}$	$3.46 \cdot 10^{-6}$	$4.77 \cdot 10^{-6}$	$1.46 \cdot 10^{-6}$
	i=6	i=7	i=8	i=9	i=10
R_i	0.0622	0.0668	0.0935	0.0631	0.0695
χ_i	0.0285	0.0305	0.0416	0.0289	0.0316
$|f_i(\bar{\mathbf{e}}_i)|$	$9.23 \cdot 10^{-6}$	$4.45 \cdot 10^{-6}$	$1.015 \cdot 10^{-5}$	$2.783 \cdot 10^{-5}$	$4.95 \cdot 10^{-6}$

The NBM algorithm as well as the SOI algorithm (available in [6]) applied to $\mathbb{X}^\varepsilon$ detect the same support of f, since they compute the polynomial $h = y^2 - 1.0041x - 2.0089y + 2.1287$ which almost vanishes at $\mathbb{X}$. Nevertheless h is not a solution to our problem: in fact, $h = 0$ (the dashed line in Figure 2) contains no admissible perturbation of $\mathbb{X}^\varepsilon$ since it does not cross the ε-neighborhood of the first point of $\mathbb{X}$.

In the following example a fairly large set of points of $\mathbb{R}^5$ is considered.

Example 5.2. Let $\mathbb{X} \subset \mathbb{R}^5$ be a set of points created by perturbing by less than 0.1 the coordinates of 25 points lying on the affine variety $\mathcal{Z}(g)$ where

$$g \;=\; x_4^2 - \tfrac{17}{41}x_4x_5 - 2x_5^2 - \tfrac{10}{41}x_1 + \tfrac{21}{41}x_2 - \tfrac{74}{41}x_3 + \tfrac{93}{41}x_4 - \tfrac{36}{41}x_5 + \tfrac{39}{41}$$

Let $\varepsilon = 0.1$, $\delta = 2\varepsilon$ and $k = 2$; the LPA applied to $\mathbb{X}^\varepsilon$ performs 11 iterations in a computational time of 24 seconds. During the 11th iteration the term $t = x_4^2$ is considered and the output $\bar{\mathbf{e}}$, f, $\mathcal{O}$ is returned, where $\|\bar{\mathbf{e}}\|_\infty = 0.0789$

$$
\begin{aligned}
f \;=\;& x_4^2 - 0.421x_4x_5 - 1.994x_5^2 - 0.244x_1 + 0.405x_2 - 1.779x_3 + 2.328x_4 \\
& -0.685x_5 + 1.124 + 10^{-2} \cdot (0.056x_1x_5 - 2.549x_2x_5 + 1.679x_3x_5) \\
\mathcal{O} \;=\;& \{1, x_5, x_4, x_3, x_2, x_1, x_5^2, x_4x_5, x_3x_5, x_2x_5, x_1x_5\}
\end{aligned}
$$

The set $\mathbb{X}(\bar{\mathbf{e}})$ is an admissible perturbation of $\mathbb{X}^\varepsilon$; we can consider the polynomial f almost vanishing at $\mathbb{X}(\bar{\mathbf{e}})$ since $\|f(\mathbb{X}(\bar{\mathbf{e}}))\|_2 / \|f\| \approx 3.42 \cdot 10^{-4}$. We apply Theorem 4.3 to this example; since for each i the inequality (4.6) is satisfied, we conclude that there exists $\mathbf{e}^*$ such that $\mathbb{X}(\mathbf{e}^*)$ is an admissible perturbation of $\mathbb{X}^\varepsilon$, $f(\mathbb{X}(\mathbf{e}^*)) = 0$ and $\|\mathbf{e}^* - \bar{\mathbf{e}}\|_\infty < 0.097$.

We observe that the minimum of the degrees of the polynomials of $\mathcal{I}(\mathbb{X})$ is 3, and this does not suggest that the points of $\mathbb{X}$ lie close to the variety $\mathcal{Z}(g)$.

On the contrary, since the coefficients of f differ only slightly from the corresponding coefficients of g, the LPA allows us to recover the "approximate" geometrical configuration of $\mathbb{X}$.

The following example (taken from [7]) is different from the standard cases deriving from the analysis of real-world phenomena. It points out that the LPA has various possibilities of applications; in this case it is employed to obtain the numerical implicitization of a parametric curve.

Example 5.3. We consider the parametric equations for a Bézier curve:

$$
\begin{aligned}
x(t) &= 4t(2t^5 - 3t^4 + 8t^2 + 6t + 3)/(t^6 - 3t^5 + 3t^4 + 3t^2 + 3t + 1) \\
y(t) &= 6t(4t^4 + 9t^3 - 9t^2 - 9t + 5)/(t^6 - 3t^5 + 3t^4 + 3t^2 + 3t + 1)
\end{aligned} \tag{5.1}
$$

whose implicit equation is $g = 0$ where

$$
\begin{aligned}
g &= x^3 - \tfrac{2}{1269}x^2 y - \tfrac{28}{423}xy^2 + \tfrac{224}{34263}y^3 - \tfrac{15712}{1269}x^2 - \tfrac{56}{1269}xy + \tfrac{848}{3807}y^2 \\
&\quad + \tfrac{44480}{1269}x - \tfrac{17792}{1269}y
\end{aligned}
$$

We consider the set of points

$$
\begin{aligned}
\mathbb{X} = \; &\{(0,0), (-1.3581, -4.7661), (2.0956, 2.0315), (4.6884, -0.3349), \\
&(-2.7205, -11.6848), (-7.2835, -40.9773), (6.7793, -1.4114), \\
&(8.6024, 1.4575), (10.5250, 8.8937), (12.6213, 19.7217)\}
\end{aligned}
$$

created by evaluating the parametric equations (5.1) at 10 random values of the parameter $t \in (-1, 2)$, and rounding off up to 10^{-4}.

Let $\varepsilon = 10^{-4}$, $\delta = 2\varepsilon$ and $k = 2$; the LPA applied to $\mathbb{X}^\varepsilon$ performs 9 iterations. During the first 8 iterations condition (4.5) is satisfied, and the set $\mathcal{O} = \{1, y, x, y^2, xy, x^2, y^3, xy^2, x^2y\}$ is constructed. During the 9-th iteration the term $t = x^3$ is considered: at step **LP4** the RFA applied to the current residual function ρ returns $\bar{\mathbf{e}}_1 \in Q_\varepsilon \cap D_{\mathcal{O}}$, therefore at step **LP5** a polynomial f_1 is computed, and the algorithm stops with the output:

$$
\begin{aligned}
\bar{\mathbf{e}}_1 &= 10^{-6} \cdot (8.13, -1.99, 0.92, -4.90, 1.40, -7.63, -3.30, 23.79, -24.90, \\
&\quad 8.50, -3.25, 0.36, -0.13, 1.11, 4.73, -7.19, 9.02, -8.41, 5.32, -1.54) \\
f_1 &= x^3 - 0.00159x^2y - 0.06618xy^2 + 0.00654y^3 - 12.3814x^2 - 0.04396xy \\
&\quad + 0.22269y^2 + 35.0514x - 14.0206y - 0.00033 \\
\mathcal{O} &= \{1, y, x, y^2, xy, x^2, y^3, xy^2, x^2y\}
\end{aligned}
$$

Note that the coefficients of f_1 and g are very close to each other; nevertheless, an even better result can be obtained by taking into account the exact information that the Bézier curve $g = 0$ vanishes at the origin. To this aim we impose that the constant term of the polynomial computed by the LPA is zero; under this assumption and using the above values for the parameters ε, δ, k, the LPA applied to $\mathbb{X}^\varepsilon$ stops with $\bar{\mathbf{e}}_2$, f_2, $\mathcal{O}$, where

$$
\begin{aligned}
\bar{\mathbf{e}}_2 \;=\; & 10^{-6} \cdot (0, -2.09, 0.96, -5.13, 1.46, -8.00, -3.45, 24.92, -26.09, 8.90, \\
& 0, 0.37, -0.14, 1.16, 4.95, -7.53, 9.44, -8.81, 5.57, -1.61) \\
f_2 \;=\; & x^3 - 0.00158x^2y - 0.06619xy^2 + 0.00653y^3 - 12.3814x^2 - 0.04407xy \\
& + 0.22271y^2 + 35.0513x - 14.0205y
\end{aligned}
$$

We observe that f_2 is better than f_1 because its coefficients provide an even closer approximation of the coefficients of g, and its support and the support of g are the same.

6 Appendix

The following proposition illustrates how to compute each column of the Jacobian matrix $J_\rho(\mathbf{e})$, that is the vector $\partial\rho(\mathbf{e})/\partial e_{ki}$, without passing through the explicit expression of the function ρ.

Proposition 6.1. *Let $\mathcal{O} = \{t_1 \ldots t_m\}$ be a finite set of terms, let t be a term, $t \notin \mathcal{O}$, and let ρ be given by (4.3); then*

$$
\frac{\partial\rho(\mathbf{e})}{\partial e_{ki}} = \left[I - M_\mathcal{O}(\mathbf{e})M_\mathcal{O}^\dagger(\mathbf{e})\right]\left[\frac{\partial t(\mathbf{e})}{\partial e_{ki}} - \frac{\partial M_\mathcal{O}(\mathbf{e})}{\partial e_{ki}}\alpha(\mathbf{e})\right] - M_\mathcal{O}^\dagger(\mathbf{e})^t\frac{\partial M_\mathcal{O}^t(\mathbf{e})}{\partial e_{ki}}\rho(\mathbf{e})
$$

where $I \in \mathrm{Mat}_s(\mathbb{R})$ is the identity matrix, $M_\mathcal{O}^\dagger(\mathbf{e}) = \left(M_\mathcal{O}^t(\mathbf{e})M_\mathcal{O}(\mathbf{e})\right)^{-1} M_\mathcal{O}^t(\mathbf{e})$, $\partial t(\mathbf{e})/\partial e_{ki}$ is the zero vector except the k-th coordinate which is equal to $\partial_i t(p_k(\mathbf{e}_k))$ and $\partial M_\mathcal{O}(\mathbf{e})/\partial e_{ki}$ is the zero matrix except the k-th row which is equal to $(\partial_i t_1(p_k(\mathbf{e}_k)) \ldots \partial_i t_m(p_k(\mathbf{e}_k)))$.

Proof. From the equality $\rho(\mathbf{e}) = t(\mathbf{e}) - M_\mathcal{O}(\mathbf{e})\alpha(\mathbf{e})$, deriving matrices and vectors componentwise w.r.t. e_{ki}, we obtain

$$
\frac{\partial\rho(\mathbf{e})}{\partial e_{ki}} = \frac{\partial t(\mathbf{e})}{\partial e_{ki}} - \frac{\partial M_\mathcal{O}(\mathbf{e})}{\partial e_{ki}}\alpha(\mathbf{e}) - M_\mathcal{O}(\mathbf{e})\frac{\partial\alpha(\mathbf{e})}{\partial e_{ki}} \tag{6.1}
$$

Deriving componentwise each side of $M_{\mathcal{O}}^t(\mathbf{e})M_{\mathcal{O}}(\mathbf{e})\alpha(\mathbf{e}) = M_{\mathcal{O}}^t(\mathbf{e})t(\mathbf{e})$ w.r.t. e_{ki} we have

$$\frac{\partial\alpha(\mathbf{e})}{\partial e_{ki}} = \left(M_{\mathcal{O}}^t(\mathbf{e})M_{\mathcal{O}}(\mathbf{e})\right)^{-1}\left[\frac{\partial M_{\mathcal{O}}^t(\mathbf{e})}{\partial e_{ki}}\rho(\mathbf{e}) + M_{\mathcal{O}}^t(\mathbf{e})\left(\frac{\partial t(\mathbf{e})}{\partial e_{ki}} - \frac{\partial M_{\mathcal{O}}(\mathbf{e})}{\partial e_{ki}}\alpha(\mathbf{e})\right)\right]$$

and the thesis follows by substituting the expressions of $\partial\alpha(\mathbf{e})/\partial e_{ki}$ in (6.1).

For describing $\frac{\partial t(\mathbf{e})}{\partial e_{ki}}$ and $\frac{\partial M_{\mathcal{O}}(\mathbf{e})}{\partial e_{ki}}$ we consider the derivative of a generic term $q = \Pi_{i=1}^n x_i^{\beta_i}$. Since $q(\mathbf{e}) = \left(\Pi_{i=1}^n(p_{1i} + e_{1i})^{\beta_i}, \ldots, \Pi_{i=1}^n(p_{si} + e_{si})^{\beta_i}\right)^t$ the vector $\frac{\partial q(\mathbf{e})}{\partial e_{ki}}$ is the zero vector except its k-th coordinate which is equal to $\partial_i q(p_k(\mathbf{e}_k))$. It follows that only the k-th coordinate of the vector $\frac{\partial t(\mathbf{e})}{\partial e_{ki}}$ and the k-th row of the matrix $\frac{\partial M_{\mathcal{O}}(\mathbf{e})}{\partial e_{ki}}$ can be different to zero: they are equal to $\partial_i t(p_k(\mathbf{e}_k))$ and to $(\partial_i t_1(p_k(\mathbf{e}_k)) \ldots \partial_i t_m(p_k(\mathbf{e}_k)))$ respectively. $\qquad\square$

Acknowledgments

References

[1] J. Abbott, C. Fassino, and M. Torrente, *Thinning Out Redundant Empirical Data.* Math. Comput. Sci. 1 (2007), no. 2, 375–392.

[2] J. Abbott, C. Fassino, and M. Torrente, *Stable Border Bases for Ideals of Points.* J. Symbolic Comput. 43 (2008), no. 12, 883–894.

[3] C. H. Bischof and G. Quintana-Ortì, *Computing Rank-Revealing QR Factorizations of Dense Matrices.* ACM Trans. Math. Software 24 (1998), no. 2, 226–253.

[4] C. H. Bischof and G. Quintana-Ortì, *Algorithm 782: Codes for Rank-Revealing QR Factorizations of Dense Matrices.* ACM Trans. Math. Software 24 (1998), no. 2, 254–257.

[5] B. Buchberger and H. M. Möller, *The construction of multivariate polynomials with preassigned zeros.* Proc. EUROCAM '82, LNCS, 144 (1982), 24–31.

[6] CoCoA Team, CoCoA: a system for doing computations in Commutative Algebra. Available at `http://cocoa.dima.unige.it/`

[7] R. M. Corless, M. W. Giesbrecht, I. S. Kotsireas, and S. M. Watt, *Numerical implicitization of parametric hypersurfaces with linear algebra.* Proc. Artificial Intelligence with Symbolic Computation, (AISC 2000), Lecture Notes in Artificial Intelligence, 1930 (2000), Springer Verlag, 174–183.

[8] B. N. Datta, *Numerical Linear Algebra and Applications.* SIAM, Philadelphia, 2010.

[9] J. W. Demmel and N. J. Higham, *Improved Error Bounds for Underdetermined System Solvers.* SIAM J. Matrix Anal. Appl. 14 (1993), no. 1, 1–14.

[10] C. Fassino, *Vanishing Ideal of Limited Precision Points.* J. Symbolic Comput. 45 (2010), no. 1, 19–37.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations. Third edition.* Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, 1996.

[12] GSL(release 1.15)- GNU Scientific Library. Available at `http://www.gnu.org/software/gsl/`

[13] D. Heldt, M. Kreuzer, S. Pokutta, and H. Poulisse, *Approximate computation of zero-dimensional polynomial ideals.* J. Symbolic Comput. 44 (2009), no. 11, 1566–1591.

[14] Y. P. Hong and C. T. Pan, *Rank-Revealing QR Factorizations and the Singular Value Decomposition.* Math. Comp. 58 (1992), no. 197, 213–232.

[15] M. Kreuzer and H. Poulisse, *Subideal border bases.* Math. Comp. 80 (2011), no. 274, 1135–1154.

[16] M. Kreuzer, L. Robbiano, *Computational Commutative Algebra 1.* Springer-Verlag, Berlin, 2000.

[17] M. Kreuzer and L. Robbiano, *Computational Commutative Algebra 2.* Springer-Verlag, Berlin, 2005.

[18] B. Mourrain, *A New Criterion for Normal Form Algorithms.* Proc. AAECC, LNCS 1719 (1999), 430–443.

[19] L. Robbiano and J. Abbott (Eds.), *Approximate Commutative Algebra.* Text & Monographs in Symbolic Computation, Springer, 2010.

[20] T. Sauer, *Approximate varieties, approximate ideals and dimension reduction.* Numer. Algor. 45 (2007), 295–313.

[21] H. J. Stetter, *Numerical Polynomial Algebra.* SIAM, Philadelphia, PA, USA, 2004.

[22] H. J. Stetter, *"Approximate Commutative Algebra" - an ill-chosen name for an important discipline.* ACM Commun. Comput. Algebra 40 (2006), no. 3.

[23] M. Torrente, *Application of algebra in the oil industry.* Ph.D. Thesis, Scuola Normale Superiore, Pisa, 2009.

[24] H. F. Walker and L. T. Watson, *Least-change secant update methods for underdetermined systems.* SIAM J. Numer. Anal. 27 (1990), no. 5, 1227–1262.